

NASA Heliophysics Science Data Management Policy

Version 1.0

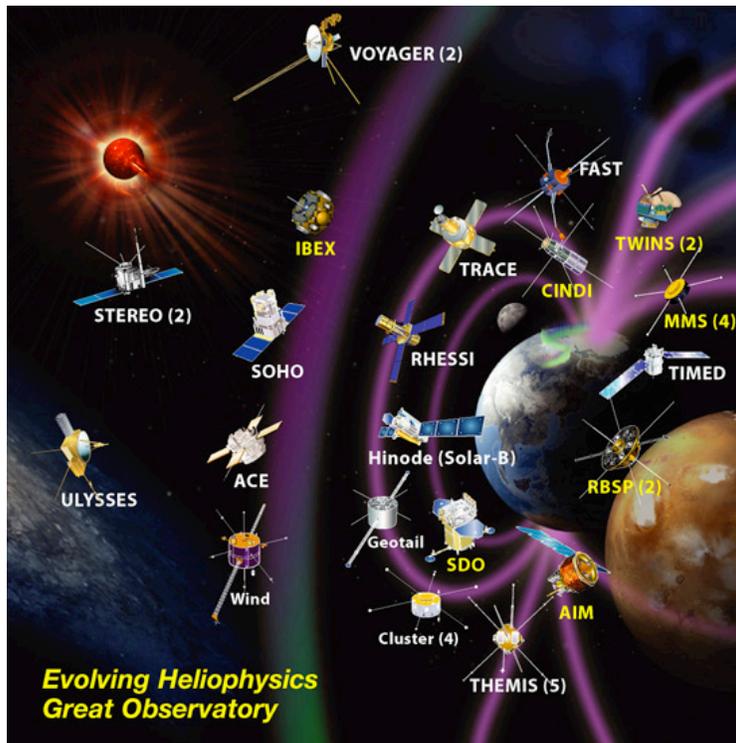
25 June 2007

Heliophysics Science Data Management Policy

Change Log

6/25/07 Release of version 1.0

Heliophysics Science Data Management Policy



To: The Heliophysics Community

The Heliophysics Science Data Management Policy is an important step forward in the evolution of the Heliophysics Data Environment that is the public face of NASA's Heliophysics Great Observatory. This policy provides a blueprint for the HPDE, tracing the data lifecycle from measurements to final archives. Our new environment, in which data are efficiently served through Virtual Observatories from distributed active and (longer-term) resident archives, should provide the infrastructure needed to achieve our scientific goals and objectives. Multi-instrument and multi-mission studies of the Sun and its effects on the heliosphere and on the magnetospheres and ionospheres of the planets will be facilitated by the approach presented here, thus enabling the attack on the next generation of HP science problems, including the understanding needed for robust space weather prediction and the related exploration of our solar system.

This data policy is vital to our research community and it incorporates the community's input throughout. This is a living document, to be modified as needed as our science program evolves. We welcome your feedback, as only through such interaction will the HPDE continue to be responsive to community needs.

Richard R. Fisher
Director, Heliophysics Division

Heliophysics Science Data Management Policy

Contents

Executive Summary

1. Introduction, Purpose, and Scope
2. The Components of the Data Environment and their Roles
3. The Mission/Data/Review Lifecycle
4. The Role of Standards: Formats, Data Model
5. Final Archiving and Continued Serving of Data
6. Plans for HP Science Data Management Policy Review and Revision

References

Appendix A: The Heliophysics Data Environment Rules of the Road

APPENDIX B: A Framework for Space and Solar Physics Virtual Observatories (Executive Summary)

APPENDIX C: A Space and Solar Physics Data Model from the SPASE Consortium (Executive Summary)

APPENDIX D: Project Data Management Plans

APPENDIX E: Mission Archive Plans

APPENDIX F: Resident Archive Functions

Figure 1: Heliophysics Mission Data Lifecycle

Heliophysics Science Data Management Policy

Executive Summary

Heliophysics (HP) research seeks to determine and model the nature and dynamical interactions of the Sun, the heliosphere, and the plasma environments of the planets based on data from a fleet of spacecraft termed the “Heliophysics Great Observatory.” Achieving the desired understanding requires easy access to data and tools from a distributed set of active archives, each of which has its own architecture and formats. This Policy document provides HP policy and guidelines for preparing, accessing, using, and archiving HP data throughout its lifetime. The basic principles for the HP Data Environment (DE) are the involvement of scientists in each stage of the process, and the acceptance of the goal of openly accessible data that are independently scientifically usable. The data environment described here is guided by a “top-down” vision provided by HQ with community input, but it is implemented from the bottom up, built from peer-reviewed data systems driven by community needs and founded on community-based standards. Consistent with this approach, data providers and data users share responsibility for the quality and proper use of the data for research.

Spaceflight projects (“missions”) are the core of the HPDE. The proposals for these define the science goals that determine the required data products. The production and serving of the data from missions is governed by the Project Data Management Plan (PDMP) until after the termination of the active phase of a mission. A Mission Archive Plan, initially formulated at the first Senior Review for an extended mission and in effect until the mission’s end, will guide the preparation of lasting data products. After a mission ends, its data will typically remain accessible through a Resident Archive that maintains easy access to data and to expertise for its use. Permanent archiving, with continued but unsupported access, is a natural endpoint in the data cycle rather than a special task.

The data from missions is to be made available by them both directly and via Virtual Observatories (VOs) that will provide one-stop access to data from many missions along with tools for cross-mission analysis and visualization. NASA HQ will provide the vision for the DE and insure, with community input from the HP Data and Computing Working Group (HPDCWG) and Senior Reviews of both missions and Data Centers, that the vision is carried out. NASA Data Centers will provide various services, including archives and cross-disciplinary access to data and services. All components of the HPDE will involve competitive selections and reviews to assure the best quality.

This document provides an overview of the components of the HPDE and their relationships, a timeline of significant events in the data lifecycle, guidelines for the preparation of Project Data Management Plans and Mission Archive Plans, guidelines for the long-term serving and archiving of data, and a plan for keeping this Data Policy updated in light of changing technology and community needs. The document is intended for all those who deal with HP data including people developing or managing missions, anyone providing or serving HP data (such as missions, virtual observatories, and data centers), those proposing for missions or data environment components, and users of the data.

Heliophysics Science Data Management Policy

1. Introduction, Purpose, and Scope

NASA's Science Mission Directorate conducts scientific exploration that is enabled by access to space [1]. The NASA Strategic Objective for Heliophysics, in particular, is to "explore the Sun-Earth system to understand the Sun and its effects on Earth, on the Solar System, and on the space environment conditions that will be experienced by explorers, and demonstrate technologies that can improve future operational systems" [2]. Our solar system is governed by the Sun through gravity, radiation, and through streams and gusts of solar wind and magnetic fields that interact with the fields and atmospheres of planetary bodies. The space weather produced by the solar effects are seen in the ozone layer; in climate change; and in effects on radio and radar transmissions, electrical power grids, and the electronics of spacecraft. HP seeks to understand how and why the Sun varies, how planetary systems respond, and how human activities are affected. As we reach beyond the confines of Earth, this science will enable the space weather predictions necessary to safeguard the outward journey of human and robotic explorers.

A key step toward the goals of the HP missions is the production and analysis of high-quality data from space platforms. This is no longer a matter of a PI team executing an experiment in space; our research goals now require an integration of data from the many instruments and missions comprising the "Heliophysics Great Observatory" as well as complimentary sources of data used to perform Heliophysics research. The success of this effort depends on having scientific involvement in all stages [3, 4] of data production, dissemination, and archiving, with a close collaboration between scientific and technical teams. Two overarching principles also essential to achieving the goals of current Heliophysics programs are:

1. Embracing NASA's open data policy that high-quality, high-resolution data, as defined by the mission goals, will be made publicly available as soon as practical, and
2. Adhering to the goal of early and continuing independent scientific data usability, which requires uniform descriptions of data products, adequate documentation, sustainable and open data formats, easy electronic access, appropriate analysis tools, and care in data preservation.

This view involves the data users from the general science community as responsible partners in the improvement of the data environment and of the data products themselves, as detailed in the HP Data Environment Rules of the Road (Appendix A). To assure responsiveness to the community and high quality, all aspects of the HPDE will involve competition through proposals or periodic reviews that include both quality assessment and consideration of plans for and the use of community feedback, and thus the HPDE is "market-driven."

The data lifecycle for HP missions envisioned here starts with the science goals for a new mission. These goals yield a set of constraints on what quantities need to be measured and at what sensitivity and resolution to answer the questions posed. The measurements are embodied in a set of data products, first articulated in the mission proposal, and later

Heliophysics Science Data Management Policy

made more precise in the PDMP that forms the basis for data gathering, reduction, and serving for the active mission. In the early phases of an active mission, data reduction routines may change continually, and the best products may only be produced for time intervals of high interest. The experience gained here leads to the ability to routinely produce high-quality, carefully documented data, served to the community using easily understood formats and delivery mechanisms. As the mission ages, and especially in the extended phase, usage typically decreases and the argument for further data refinement becomes less compelling. At this stage, a Mission Archive Plan will guide the preparation of a final set of best products and documentation that will continue to serve the community well beyond the end of the mission. The maintenance and serving of the mission archive can be performed by a “Resident Archive” (RA) associated with mission data experts. At some point the support of the datasets through an RA may no longer be cost effective, but the data product files will still be useful as served from a permanent archive. The final transition should be quite simple when the other steps in the process are performed well.

Technological advances and increasing data volumes lead naturally to having a distributed data environment, with many remote data archives at provider and other archive sites linked through software services known as Virtual Observatories (VOs). The VOs do not replace the primary repositories, but they enhance the ability to obtain and use data efficiently across a broad range of observatories, instruments, and data formats. VOs will allow scientists to access data from many missions from one Web location, or even from their own software applications directly.

The use of internet-based services founded on community-evolved practices and standards allows for a “bottom-up” implementation of specific, peer-reviewed VOs and other services all working toward a “top-down” vision provided by NASA HQ with community input. To be effective, the VO groups need to work together and with the data providers and users to establish and maintain standards that allow effective communication and interoperability. Traditionally, collaboration with other nations and agencies has fostered the inclusion of essential non-NASA holdings in the HP Data Environment, and NASA HP will continue to encourage such interactions.

The purpose of this document is to present the HP Science Data Management Policy, which is based on the above principles and overview. This policy describes the philosophy for science data management for the science programs sponsored by NASA’s HP Division, and its scope encompasses all phases of the mission and data life cycles. In particular, it provides:

- guidelines for the data system aspects of AOs and proposals for missions and instruments,
- recommendations for Project Data Management Plans (PDMPs) and Mission Archive Plans (MAPs),
- the functions of Resident Archives,

Heliophysics Science Data Management Policy

- the role of the National Space Science Data Center (NSSDC) as the manager of the permanent archives of HP data.
- a reference for senior review proposals and follow ups,
- guidelines for maintaining data availability in the long-term archive phase,
- a reference for NRAs for enhancing the HP data environment, and
- a measuring stick for peer reviews related to the data environment

2. The Components of the Data Environment and their Roles

The HPDE is that portion of NASA's HP Division concerned with the processing, serving, storing, and preserving of data and associated data analysis tools from HP missions and the broader HP community. It will increasingly involve the "data" produced by models as well, which is also essential to the HP goals outlined above, but this will not be the focus of this document. This section provides an overview of the relevant HP components and their role in the data environment.

2.1 Heliophysics Division Overview

As stated in the Heliophysics Roadmap [2], the HP endeavor relies on a number of major mission programs. The Solar Terrestrial Probes focus primarily on fundamental science questions. Living With a Star missions and partnerships target knowledge of processes that directly affect life and society. The flexible Explorer program provides an efficient means of achieving urgent strategic goals that is responsive to new knowledge, technology and priorities. Challenging Flagship and Partnership missions address important goals that cannot be funded in the baseline program. The New Millennium Program provides essential tests of new technologies using low-cost spaceflight missions. The Heliophysics Great Observatory coordinates new and existing mission elements to confront broader problems. Complementary to the missions, scientific research programs provide data analysis and theory to understand the data from the missions as well as rapid access to space measurements through the rocket and balloon programs.

What follows spells out the specific components of the HPDE and their roles and responsibilities in enabling the data flowing from the above missions to be used effectively for scientific research.

2.2 Spaceflight Projects

Spaceflight projects ("missions") are the core of the HPDE, as the providers of the data that lead to understanding. The missions should manage data so as to facilitate achievement of the mission's science goals. For the success of the DE, missions must provide open community access to the products required to attain the science goals of the proposed mission as soon as practicable, as well as to higher-level data products. This should be through an active archive of the mission's design, made to be compatible with the overall HPDE framework.

Heliophysics Science Data Management Policy

Each mission is expected to manage its own data and systems; this may be done using their own “in house” systems, or in collaboration with other data systems or centers. The creation and maintenance of the data archive will be the responsibility of each mission during the prime mission phase (Phase E). The assignment of maintenance responsibilities to the mission elements (including instrument teams) will be chosen to promote efficient processing and distribution of science data as a means of meeting the mission’s Level-1 requirements. These assignments are made early in the mission’s development and documented in the Project Data Management Plan PDMP (see below). Missions need to create and provide access to supporting material (documentation, software) required to ensure independent data usability. Missions should adhere to such data and metadata standards as practical. (See the note on standards below.)

These standards can evolve over the life of the mission. Furthermore, the uses of the data evolve as the mission matures. The underlying information technology that hosts a mission’s data environment evolves in ways not envisioned during a mission’s development phases. Capturing the evolved mission data system will be carried out by creating, in time for the first Senior Review for an extended mission, a Mission Archive Plan (MAP) and adhering to it thereafter. The MAP focuses on the content of the data and metadata files at the end the mission. The MAP will depict the status of the mission’s science data (science quality, documentation, formats, standards, and essential data analysis tools) in the final mission archive. The MAP should show the path to creating the mission’s Resident Archive(s). The MAP will be updated as the mission progresses into and through its extended mission phase.

2.3 Virtual Observatories

A recent development in the HPDE is the introduction of Virtual Observatories (VOs). The idea of VOs started as a “Digital Sky” project intended to give astronomers searchable, virtual (electronic) access to all observations of any region in all wavelengths through the use of the Internet to reach archives of data located worldwide. This project became the National Virtual Observatory in the US, and the International Virtual Observatory Alliance worldwide. Solar and space physics are now embarking on a similar unification of access to the many HP Great Observatory spacecraft and ground-based instruments. The HP VOs will primarily be intended to provide simple, uniform access to data from distributed, heterogeneous sources, but they will also enable services, such as visualization or format translation, that enhance the use of these data.

Accomplishing this unification requires a coordinated effort to link data and service providers to scientific users through software that uses nearly universal language descriptions to give a uniform face to an underlying heterogeneous and distributed set of resources. The software services that accomplish this task of linking users to data and services have now become generally known as VOs. The VOs do not typically hold data, but constructing VOs requires strong interaction with data providers to achieve the desired VO goal of uniform descriptions of data products and seamless access to them. The access to products may involve the development of visualization and query services beyond those needed for basic browse and access, but the primary goal is to make the data from many sources easily found and available in convenient form for scientific use.

Heliophysics Science Data Management Policy

The HP VOs are being initiated through calls in the annual ROSES solicitations for discipline specific VxOs (e.g., x = M for Magnetospheric or S for Solar); these proposals define the scope of each specific effort. Having VxOs will allow subfields to organize their data and approach in a way that best suits that field. NASA HQ is promoting and guiding the overall integration of the VxOs. There is now a community-wide data model (the “SPASE” Data Model; see below and Appendix C) as the essential set of uniform terms needed to ensure interoperability. The VxOs are expected to work together and with the other elements of the HPDE. Conversely, new NASA missions should work with the VxOs to promote the distribution of their data. Appendix B presents the Executive Summary of a community-based defining document for VOs, along with a link to the full document.

2.4 Resident Archives

The concept of resident archives recognizes that the data from a mission are often of great current interest long after the mission has ceased actively collecting data. Formalizing this post-mission phase will also allow for a smoother transition to a useful permanent archiving of data products. In particular, the role of the RA will be primarily to securely hold and serve data effectively, providing support to the community for the data use. This “component” of the HPDE is not really separate from the others, in that the team involved would usually involve a subset of the original mission team. RAs are described in more detail in Appendix E. The HP RA concept parallels to some extent the Planetary Data System (PDS) “data nodes” and Astrophysical Science Archive Research Centers (SARCs) but with an emphasis on keeping data near its producers, and without a preset determination of where particular datasets will reside. The serving of datasets from one RA that cross multiple missions—united, for example, by measurement type, common team members, or subfield—will be a natural part of the HPDE, depending on what is most effective for the particular data.

2.4 NASA Data Centers

The NASA Data Centers form an important aspect of the HPDE. These consist of the Solar Data Analysis Center (SDAC), the Space Physics Data Facility (SPDF), and the National Space Science Data Center (NSSDC). The latter has the responsibility for assuring that NASA data are preserved over the long term, and their role includes archiving data from Planetary Science and Astrophysics missions in addition to HP missions. For HP missions, this role will in many cases be changing from one of physically gathering and storing data to one of providing assurance that the data are preserved through the management of RAs. Part of this is expected to involve knowing what data are and should be available across the HP disciplines, based on PDMPs, VxO registries, and other input; this information should be made readily accessible. As the process of serving data during active missions becomes more uniform and better documented with the help of VOs and related activity, the problem of archiving for the long-term should become relatively easy. Resident Archives may hold data for longer times than has conventionally been the case with inactive missions, but if the RA becomes no longer cost effective, then the physical data will pass to a facility determined in collaboration with the NSSDC.

Heliophysics Science Data Management Policy

The SDAC and SPDF, which traditionally have been active data repositories, have increasingly become centers for excellence in providing multi-project, cross-disciplinary access to data and tools to support the broad range of science possible with the “Heliophysics Great Observatory.” They also produce and maintain such things as the Common Data Format (CDF) software for making and using files in a self-documenting format and the SolarSoft set of routines for solar data analysis. The role of these two centers in ingesting data directly will be decided in the future on a case by case basis, in collaboration with particular projects that may find it beneficial to use the existing expertise to solve data storage and serving problems. Among other similar roles, the SPDF functions as the active archive, effectively acting as a Resident Archive in many cases, for Space Physics Data held or to be held at the NSSDC. SDAC will serve data from some of the new solar missions in addition to serving, e.g., much of the SOHO image data; it, too, may ultimately take on a Resident Archive role in such cases.

Likewise, SDAC and SPDF can assist in the long-term maintenance of the VxO infrastructure that is now just emerging. In the solar case the core functions of VSO, the first of the VxOs, is now funded as a part of the SDAC functions. The VxOs, selected originally via the competed programs of ROSES, may need to be sustained in the longer term by a periodically reviewed arrangement, yet to be determined. One example of such a transition is the transfer of basic support of the Virtual Space Physics Observatory to SPDF. The model for other VOs has yet to be worked out, but HQ will assure that VxO functions will persist beyond the current startup grants.

It is useful to note that many other data centers exist that provide data relevant to HP science goals. In the US, the PDS, for example, provides access to NASA Planetary data; the NGDC provides a link to essential NOAA data; and CEDARWeb provides key ground-based data from NSF. Worldwide there are data centers associated with ESA (e.g., the CDPP in France) in Europe, the CSA and related efforts such as GAIA in Canada, and JAXA (e.g., the DARTS active archive) in Japan. There are virtual observatory efforts such as the European EGSO and other projects underway that will link VOs and these data centers to each other to produce a world-wide analogue of NASA’s HP Great Observatory, and the HPDE is working to make this larger goal a reality.

2.5 NASA HQ, HQ Program Offices, and Program Scientist

The role of NASA HQ has been mentioned above in various places, and obviously includes approving and overseeing missions and proposals. NASA HQ will continue to develop, with input from community groups, the overall philosophy and direction of the HPDE, as expressed in this Data Policy document. HQ will also do all it can, subject to available resources, to support the HPDE, ensuring an architecture through which data and supporting material will be community-accessible and preserved for the long term, and that will evolve as technologies and requirements evolve. HQ will be responsible for convening and using the results from Senior Reviews and NRA reviews to establish projects and priorities for the HPDE.

Heliophysics Science Data Management Policy

The NASA Program Offices oversee the design and implementation of missions, and thus are the primary point of contact between the missions and HQ. Each Program Office (e.g., LWS or STP) allocates budgets and oversees contracts with the projects that make up the program. The contracts include, in addition to the primary hardware deliverables and related measurement requirements, the requirements for data provision, access, and delivery. This Data Policy is designed to provide guidelines consistent with the contractual requirements, but provides, in addition, recommendations designed to lead to the best return on the investment in the data.

HPDE program integration is currently being facilitated by an HPDE Program Scientist who oversees the competition for DE components such as VxOs, RAs, and related tools, as well as the progress of the resulting selected projects. The PS maintains an HPDE Web site with overviews of the DE, current events of interest, and links to the HPDE component activities. He reports to and gets input from the HPDCWG on the directions of the HPDE, including on the definition and maintenance of standards such as a Data Model and formats; ensures that VxOs, data repositories, and other interested parties meet and otherwise interact frequently enough to assure an integrated data environment; and works with other nations and agencies to assure the HPDE includes all components needed for success. In the longer term, the role of the HPDE Program Scientist may be changed or replaced as the structure of the DE evolves.

2.6 HP Data and Computing Working Group

The HP Data and Computing Working Group is the principle community group discussing the HPDE. (The HPDCWG also addresses high-performance computing for HP, but this is not dealt with here). In light of their assessment of community needs, the HPDCWG will review this Data Policy; hear presentations of and comment on PDMPs and MAPs; review the progress of components of the DE, such as VxOs, as needed; consider the content of calls for DE-related SR&T and TR&T related work; and provide findings to help the HPDE be responsive to the community. The HPDCWG reports findings to the HPDE Program Managers and Scientist. The Chair of the HPDCWG reports to the Heliophysics Science Subcommittee.

2.7 Heliophysics Community: NRA-based work

The general community of Heliophysics researchers is naturally involved in the HPDE as the user of data and services, and thus it is the ultimate arbiter of what works or needs improvement. This will be true for VxOs, where the different initial interfaces and services will thrive or decline based on use, and each VO can learn from the others. In addition, the open data policy makes the community into a collective source of validation and verification of data quality. While it is the responsibility of the instrument teams to produce high quality data, the users of the data will be essential critics, and research projects where reputations are at stake provide strong motivation for correctness. While some users will use data assuming it is all valid, others will be appropriately wary and will be one of the best sources of data quality improvement, as has been found by many teams who have operated this way. The data quality and validity will also be addressed by the community as part of mission Senior Reviews and RA reviews.

Heliophysics Science Data Management Policy

In addition to fulfilling their role of maintaining data quality, community members may also help by providing specific products or services. All aspects of the HPDE involve elements of competition that maintains the quality of the DE, but certain aspects of the HPDE will be competed through the standard NRA process (ROSES), which facilitates the broadest possible involvement. The NRA-initiated work will include the establishment of VxOs and initial RAs, the development of value-added services (e.g, data mining services), the construction of general data analysis and visualization tools, and the restoration of datasets or their preparation for easy Internet access.

When tools or services prove to be of general value to the community, they can be transitioned to the support by Data Centers or other long-term means. The SPDF and the SDAC may serve, in some cases, as the mechanism of this long-term support. The Senior Reviews for the Data and Modeling Centers will review the appropriateness and effectiveness of each of the ongoing elements of the HPDE funded in this way.

3. The Mission/Data/Review Lifecycle

The data lifecycle of a mission is shown in Fig. 1. At the inception of new scientific mission, the goals of the investigations are laid out. From this, the concepts for the scientific instruments and the mission operations scenarios are developed. These steps lead directly to the specifications of the instruments and the data type, sensitivity, and resolution requirements. The project's Data Management Plan (PDMP) captures the architecture and implementation of the processing and distribution of mission data. [See Appendix D.]

The schedule of significant data-related events in the life cycle of a spacecraft project is as follows:

- Mission Proposal or Science Definition Report: An initial specification of the mission defines the science investigations to be conducted and the concepts for the scientific instruments needed to conduct these investigations.
- Preliminary Design Review: A draft Project Data Management Plan will state the mission's data requirements and how the mission's data system will be implemented in order to meet the Mission's requirements. One of requirements on the mission data system will be the production and open distribution of independently useable data. Among other things, the PMDP should include how the mission data team will provide links to the larger data environment, including VOs, and long term archives.
- Critical Design Review: The PDMP should be finalized and this document will guide the data activities of the mission through Phase E. The PDMP will be reviewed by the HQ Program Scientist along with the HPDCWG.
- Launch: The mission's data system as laid out in the PDMP will be operational and ready for production and distribution of the mission's science data once initial testing and calibration have been carried out.

Heliophysics Science Data Management Policy

- First Senior Review (beginning of extended mission): The mission will submit its initial Mission Archive Plan (MAP) that will indicate the steps that the mission needs to implement its science data archive. (See Appendix E.)
- During the Extended mission: The MAP will be updated for each subsequent senior review.
- The end of the mission operations: When data ceases to flow from a mission, either due to spacecraft failure or by direction of NASA HQ, NASA HQ will negotiate with the mission for the preparation of final products in accordance with the PDMP or MAP (whichever is current). At this stage, final, calibrated, well-documented products should be made, and while serving data from lower-level products processed on-the-fly or by the researchers themselves can continue, a set of data files and electronic documentation representing the best efforts of the mission should be produced and made available in a sustainable, easily used format.
- Termination of the mission: The steps in the MAP should be complete and the data should be ready to transition to a resident archive, multimission data center, or other depository under auspices of the NSSDC. Every two years to four years, HQ will more formally peer review the status of RAs, multimission data centers, and the NSSDC and adjust funding and directions. The RAs continue to hold and serve the data, and to provide expert assistance as needed to the community.
- RA termination: Upon the decision of an RA to not seek further funding, or upon peer review determination of the lack of further utility of the RA for particular or all of its data products, the data may pass either to another RA or into a permanent archive phase with continued serving through, e.g., a NASA Data Center as the effective RA. The holding and stewardship of the data in the permanent archive phase will be determined after consultation with the NSSDC (if it is not the sole holder), which has responsibility for the data at that point.

NASA HQ convenes, usually at two-to-four year intervals, senior reviews for HP missions and for its Data and Modeling Centers. These have become distinct review processes, as the nature of the activity is different in the two cases. These reviews provide community input through a panel selected for its relevant expertise. Based on these reviews, other NASA priorities, and the realities of the funding situation, HQ enters into contracts with each of the missions or data centers.

Other reviews may be needed if the above processes do not provide sufficient oversight of RAs and VxOs. Reviews of these activities, or subsets of them, will be convened by HQ as needed.

4. The Role of Standards; Formats, Data Model

The most important “standard” for the HPDE is a standard of behavior, namely, the acceptance of the need for open, independently useable data. In an era of distributed datasets and heterogeneous infrastructures for different missions, it will also be essential

Heliophysics Science Data Management Policy

that each part of the DE be committed to working together with the other parts. Competitive proposals for DE components, and reviews of the components, should strongly take into account the degree to which an effort takes a collaborative view and engages the community in making improvements.

The HPDE will benefit greatly from more conventional standards, but experience has shown that if these are imposed by bodies without community input they tend to be ignored. Thus the standards adopted in this DE will be based on utility as determined by actual implementation. In many areas, such as the communication between VxOs or between VxOs and repositories, the standards are negotiated within context of their use.

The formats for processing and storing of data by a PI team are prescribed to meet mission needs. Files for distribution to the scientific community at large should employ a common, supported, easily-used format. There are a number of data formats in common use, such as HDF-5 (primarily in Earth Science; netCDF is now related to this), FITS (e.g., in Astronomy and Solar Physics), CDF (increasingly common in Space Physics), and various forms of ASCII files with headers and/or independent documentation. XML-based formats may arise, such as used in the VOTables of the astronomical National Virtual Observatory and elsewhere, although none of these have been used as yet for large NASA HP data archives.

The terms used in self-documenting files or even internal to a VxO need not be standardized, but mappings to the SPASE Data Model terminology for use by VOs and other services should be provided, usually with the help of an appropriate VxO. The SPASE data model has been the result of a broad consortium of space and solar physics scientists and technologists, and has been agreed upon by the initial VxOs as a language for interoperability. Similar models have been in existence for a long time for the PDS and more recently for the NVO. The maintenance of this community-based standard will be through the existing consortium (see www.spase-group.org), which has open membership and welcomes input. The consortium now has an official release and mechanisms for timely updates. Modest funding to support SPASE efforts was initially provided through an NRA grant, but will be continuing as a part of the NSSDC budget.

5. Final Archiving and Continued Serving of Data

As mentioned above, the NSSDC manages permanent archiving facilities for HP (and other) data. Older data are useful for long-term studies and for unique characteristics such as specific instrumentation or regions sampled. Thus, it is useful to have not just a place for data to be preserved, but also to be served. The NSSDC has responsibility for assuring long-term data preservation and distribution. The NSSDC will ensure the maintenance of the permanent archive; the physical arrangements for such storage will be made in whatever manner is most economical, secure, and accessible.

Each mission will consult with the NSSDC concerning the final archiving of mission data. The path can include the holding of mission data in a Resident Archive, but eventually the data may only be served by a data facility under the supervision of or in coordination with the NSSDC.

Heliophysics Science Data Management Policy

The full science potential data of a mission, not irreversibly transformed, should be archived along with tools for its reduction to science products and documented algorithms for this process. Relevant engineering and “housekeeping” data should also be preserved. However, much more important is the archiving of the calibrated, useable best products from a mission and the associated documentation. Long-term archiving and serving of data cannot be based on serving products using on-the-fly data reduction. In general, but especially for long-term and nonspecialist use, it is desirable to have data products that are ready-for-use, and thus despiked, corrected for backgrounds, etc., and not dependent on specialized software packages. Lower level products and the software and algorithms to use them should be archived, but these become increasingly difficult to use. Other products, such as browse plots and event lists, provide increased utility and their archiving is encouraged.

6. Plans for HP Science Data Management Policy Review and Revision

This Policy document should be posted publicly and be reviewed on the same timescale as the Data and Modeling Centers, or as deemed necessary by HQ in consultation with the HPDCWG. At such times the proposed revisions should be submitted for comment and suggestions to the HPDCWG, the HP missions, and to the HP community at large, including partners from other organizations. The final decision on changes rests with HQ management.

Heliophysics Science Data Management Policy

References

1. NASA Science Plan 2007, available at <http://science.hq.nasa.gov/strategy/index.html>; direct link http://science.hq.nasa.gov/strategy/Science_Plan_07.pdf
2. Heliophysics Roadmap available at http://sec.gsfc.nasa.gov/sec_roadmap.htm; direct link http://sec.gsfc.nasa.gov/Roadmap_FINALscr.pdf
3. *Data Management and Computation Volume 1: Issues and Recommendations*, Committee on Data Management and Computation (CODMAC), R. Bernstein, et. al., Nation Academy Press, 1982.
4. *Report and Recommendations of the LWS Science Data System Planning Team*, D. G. Sibeck and T. Kucera, January 2002. Available at <http://hpde.gsfc.nasa.gov>; direct link http://hpde.gsfc.nasa.gov/LWS_Data_System_Final.html

The Heliophysics Data Environment website (<http://hpde.gsfc.nasa.gov>) provides a great deal of background on the data environment, with descriptions of recent activities, and annotated links to significant documents and to VOs.

Heliophysics Science Data Management Policy

APPENDIX A: The Heliophysics Data Environment “Rules of the Road”

(In what follows, “PI” may actually be an instrument lead in the case of PI-class missions.)

1. The Principal Investigators (PI) shall, in a timely manner, make available to the science data user community (Users) data and access methods to reach the scientifically useful data and provide analysis tools equivalent to the level that the PI uses.
2. The PI shall make available appropriate data products to the public that assist the PI’s EPO responsibilities.
3. The PI shall assure all scientifically important data and supporting material are archived to ensure long-term accessibility of the data and their correct and independent usability.
4. The PI or the appropriate VxO shall inform Users of updates to processing software and calibrations via metadata and other appropriate documentation.
5. Users should consult with the PI to ensure that the Users are accessing the most recent available versions of the data and analysis routines. VxOs should facilitate this, serving as the contact point between PI and users in most cases.
6. Browse products are not intended for science analysis or publication and should not be used for those purposes without consent of the PI.
7. Users should acknowledge the sources of data used in all publications, presentations, and reports. In some journals, this can now be done through formal citation of the data product in the reference list.
8. Users are encouraged to provide the PI a copy of each manuscript that uses the PI’s data upon submission of that manuscript for consideration of publication. On publication the citation should be transmitted to the PI and any other providers of data. (The community needs to work to find ways to make this easy/automatic.)
9. Users are encouraged to make tools of general utility and/or value added data products widely available to the community. Users are encouraged to notify the PI of such utilities or products. The User should also clearly label the product as being different from the original PI-produced data product.

Heliophysics Science Data Management Policy

APPENDIX B: A Framework for Space and Solar Physics Virtual Observatories

Results from a Community Workshop sponsored by NASA's Living With a Star Program, 27-29 October 2004

(Complete document at: http://hpde.gsfc.nasa.gov/VO_Framework_7_Jan_05.doc)

Executive Summary

The new challenges in solar and space physics, including linking solar phenomena to human consequences as studied in NASA's Living With a Star program, will require unprecedented integration of data and models across many missions, data centers, agencies, and countries. Accomplishing this requires a coordinated effort to link data and service providers to scientific users through software that uses nearly universal language descriptions to give a uniform face to an underlying heterogeneous and distributed set of resources. Such three-part entities—front-end software linked to repositories and services through "gateways" or "brokers"—represent a generalization of the ideas behind the "virtual observatory" (VO) intended to give astronomers virtual access to all observations of the sky. This workshop, held in Greenbelt, MD on 27-29 October 2004, brought together nearly 100 space and solar physicists and technologists, along with Earth scientists and astronomers, to come to basic agreements on how to proceed to build a robust data environment for future space and solar physics research based on the virtual observatory paradigm. Some of the main ideas had been in the community by other names for over a decade, but new Internet connectivity, greater emphasis on global problems to be solved with multiple spacecraft and models, and increased support by agencies has brought us to a point where the need and means are clearer for realizing an integrated data environment.

The workshop consisted of a set of plenary talks (available on a link at <http://hpde.gsfc.nasa.gov>, which also includes many presented posters and other background) that gave an overview of current efforts and issues, followed by 1-1/2 days of working groups and plenary sessions designed to clarify and elaborate the vision and plans. The above three-part VO structure was followed by the existing VOs, although the details differed. There are beginnings of integration of the current efforts, and the connections are becoming more direct. The workshop agreed on the need for agreement on at least a core of common "data model" terms, such as presented by SPASE and EGSO, although all agreed that specific communities, represented by "VxOs" ("x" being the community), would have some terms specific to their needs. Data models are much farther along in describing data products than services. The roles of the resource providers and the VOs were delineated at the workshop, with VxOs being mainly responsible for uniformity of access across providers and for higher level services and the providers for basic data quality and access, although the VO data environment should provide considerable flexibility in what tasks are performed by which parts of the system. It was agreed that a level of metadata management services, generally invisible to users, would be essential. Science services, such as format and coordinate translations,

Heliophysics Science Data Management Policy

visualization, and higher-order queries, were seen as highly desirable but not part of the central core services which consisted of finding and accessing resources in uniform ways.

It was agreed that the current VO groups should continue to coordinate their efforts. In the short term this will be on an informal basis, but longer term there should be a coordinating group consisting of VO representatives (scientific and technical) and users from the scientific and broader community. Initially these may be agency specific groups, but interagency and international coordination, which is a natural outgrowth of much current work, will be needed and should continue to be part of workshops and other efforts. While the data environment is becoming established, data and service providers can be describing their resources in uniform data model terms and providing feedback to groups working on the data models; making data and services machine-accessible with APIs or other means as current resources allow; and linking to current VOs or making VxO alliances. In addition to continuing to coordinate their efforts, the VOs should seek community feedback on current VO interfaces and other issues.

Heliophysics Science Data Management Policy

APPENDIX C: A Space and Solar Physics Data Model from the SPASE Consortium

(Complete document at: http://www.spase-group.org/data/doc/spase-1_1_0.pdf)

Executive Summary

The Solar and Space Physics communities need improved methods and procedures to facilitate finding, retrieving, formatting, and obtaining basic information about data essential for their research. With the increasing requirement for data from multiple sources, this need has become increasingly important. It has been long recognized that a uniform method to describe data and other resources is the key to taking the next steps in improving the data environment. The SPASE (Space Physics Archive Search and Extract) Data Model provides a basic set of terms and values organized in a simple and homogeneous way, to facilitate access to Solar and Space Physics resources. The SPASE Data Model is comparable to the data models developed by the Planetary Data System (PDS) and the International Virtual Observatory Alliance (IVOA) for planetary and astronomical data, respectively. The SPASE Model will provide the detailed information at the parameter level required for Solar and Space Physics applications.

The SPASE consortium is an international team of space and solar physicists and information scientists. It first examined many existing data models, but found none to be adequate. A set of terms based on a half-dozen or so of the most complete of such models was refined based on applying the model at various levels of detail to a large number of existing products to arrive at the current version. The major creators of SPASE-based product descriptions are expected to be individual data and model providers, data centers and domain-based Virtual Observatories (“VxOs”). The SPASE Data Model will continue to evolve in a controlled way as data and service providers and benefiting researchers suggest improvements to extend its framework of common standards. Success of the model will be measured by the extent of community support and use.

The present Data Model provides enough detail to allow a scientist to understand the content of Data Products (e.g., a set of files for 3 second resolution Geotail magnetic field data for 1992 to 2005), together with essential retrieval and contact information. A typical use would be to have a collection of descriptions stored in one or more related internet-based registries of products; these could be queried with specifically designed search engines which link users to the data they need.

The Data Model also provides constructs for describing components of a data delivery system. This includes repositories, registries and services. This document provides potential users of SPASE with the Data Model for review and use. The document has an overview of the origins and the concepts of the data model, and presents the set of elements in a hierarchy that shows the natural relationships among them. Also included are usage suggestion, pedagogic examples, and a complete set of definitions of terms and enumerated lists.

Heliophysics Science Data Management Policy

APPENDIX D: Project Data Management Plans

The Project Data Management Plan (PDMP) is the interface document between NASA, the mission systems, and the instrument teams that describes the science and ancillary data associated with the mission and how the data will be managed. This document describes how the mission will meet the Level-1 requirements that address the preparation and distribution of processed science data for the general community.

The science teams (instrument providers and Project Scientists) for each mission will develop a PDMP that defines the data, processing approach and implementation, data and documentation products, data availability, and storage and archival strategies. It will also define the access method(s) for the HP scientific community.

Signers of these documents will include the Project Manager, Project Scientist, and each Principal Investigator or Instrument Lead. Others may also need to sign this plan, depending on the Project-specific situation.

Typically, the PDMP will be available in draft form at the time of Preliminary Design Review for the mission, and signed at the time of Critical Design Review. The PDMP may be revised from time to time, but it will be augmented and eventually superseded by a Mission Archive Plan. The HPDCWG will generally review and comment on the PDMP.

Each data provider will be expected to generate and make available metadata and other supporting material on the data products, spacecraft, and instrumentation appropriate to their investigation. The details of these will be defined during discussions with the Project and Program personnel during the drafting of the PDMP. The intent of such metadata and materials will be to make the data correctly and independently useable for science investigations.

Each PDMP will:

- Provide and forecast for the mission's long-term archiving of the data and tools. The PDMP will recognize that the final mission archive will be specified in its Mission Archive Plan to be submitted to HQ for approval prior to the termination of the mission, typically in the first Senior Review proposal for an extended mission. (See Appendix E).
- State how that Project will support the broader HP goals of cross-disciplinary science, to include the use of those data by interdisciplinary scientists. Typically, this will involve collaboration with an appropriate Virtual Observatory.
- State how the ancillary and anecdotal data relevant to the mission will be captured, archived, and made available as part of the mission dataset.
- State how the mission data will interface to the HP data environment that is anticipated to exist at the time of launch.

Heliophysics Science Data Management Policy

Examples of information that are appropriate for each data provider to include in a PDMP are:

- Data flow from telemetry to science level products
- Specifications of data (including levels as defined by the mission) and estimates of data volume and frequency
- Proposed data distribution capacity
- Identification of data that are made available to the public
- Description of the means that data are made available to the public
- Schedule for making these data available
- Definitive orbit and attitude data disposition (generation, capture, distribution and storage).
- Engineering telemetry disposition (e.g., capture, distribution, archive)
- Calibration data disposition
- Description of documentation to be provided on datasets, instruments, and spacecraft relevant to data usability
- Metadata schemes to be employed, and the relation to the SPASE Data Model
- Data format specification or references (e.g., FITS, CDF, HDF, Documented ASCII, etc.)
- Processing and analysis tools to be made available, and the method for this
- Reprocessing strategy, if appropriate
- Catalogues of data or events that will be produced
- Technical support that will be provided for data use
- Back-up strategy to be implemented (routine and catastrophic)
- Plan for long-term data serving and preservation

Heliophysics Science Data Management Policy

APPENDIX E: Mission Archive Plan

A Mission Archive Plan (MAP) will be prepared by a mission team before it enters into its extended phase. The purpose of the plan is to lay out those steps needed to be completed by the mission team to ensure that the appropriate mission data archives have been prepared prior to the termination of the mission. The plan will be able address advances in Information Technology that have occurred since preparation of the PDMP and the development of its data system. Also the plan will be able to adopt new developments in the architecture of HPDE.

The plan will describe the current state of the mission's scientifically relevant data products and describe the steps needed to complete the mission archive, including the final list of products. The MAP should contain a roadmap for creating or using existing resident archives of mission data in the post-operations phase. The implementation of the MAP will be completed prior to planned termination of the mission or soon as possible after an unplanned termination of the mission. The MAP will be submitted as part of a mission's proposal to the senior reviews of the Heliophysics operating missions. Once reviewed by the senior review, the subsequent oversight of the implementation of the plan will be made by the mission's project scientist and Mission PI (if applicable). The plan will be updated periodically during the extended phase.

The MAP should include:

- An assessment of the status of existing data, ephemeris, attitude, engineering, and any other (e.g. browse, higher-level, event list, or combined) products, and of the documentation associated with the production and validation of these.
- An assessment of the status of the relevant documentation of the spacecraft, instruments, and instrument calibrations.
- A realistic plan and schedule for producing a set of final data and ancillary products (not "level zero plus software"), with a complete list of these products and their formats. Any provision of other than calibrated, highest resolution products (in addition to possible lower-resolution or higher level products) should be justified.
- A listing of the documentation to be provided on all products, instruments, and calibrations, and a plan for providing these to users such that they will be able to assess the utility of the scientific data. The relationship of metadata to the SPASE data model should be discussed.
- A listing of all analysis tools to be provided to the community, and details of how they are to be served.
- Details of how the data are to be served, including through VOs, and how this serving can be maintained for the long term through Resident and/or Permanent Archives.

Heliophysics Science Data Management Policy

APPENDIX F: Resident Archive functions

A resident archive (RA) will be created by a once-active mission to continue to serve mission data or a subset of a mission's data (e.g. data products for a single instrument) after the mission has ended. This arrangement is intended to keep those most familiar with the data and its caveats involved such that a user will have access to expert assistance in using the data for research. Typically, a mission PI or instrument lead would be the PI on an RA proposal, but there is no restriction on who can apply or on possible arrangements with other RAs or data centers.

The prioritized functions of the resident archive are as follows:

1. Ensure that the mission data are served to the general community in an efficient and scientifically useful manner consistent with the community data environment guidelines (including as stated here and elsewhere in this Data Policy);
2. Maintain the integrity of the data by safeguarding against data loss; this could be achieved by the use of mirror sites (e.g., the NSSDC, etc.), as well as with such tools as checksums;
3. Provide expert assistance with data issues;
4. Preserve and serve documentation of the data (including mission, instrument, and PI information) as required to maintain independent usability;
5. Obtain community feedback to ensure success;
6. Make sure the data will be archived and available after the RA is no longer deemed useful or cost effective (e.g., transfer to another RA or a long-term repository by agreement with the NSSDC).

The data services provided by the resident archive must be coordinated with the HP virtual observatories and be vital to on-going Heliophysics research activities or important to future research activities.

The data services to be maintained include the open, electronic distribution of RA data, the serving of the metadata for the RA data sets, and the provision of documentation describing the resident data including calibration and validation procedures and methodologies. The maintenance of RAs includes ensuring adequate security is provided to preclude the irreversible loss of mission data. The RA is expected to provide user support for both virtual observatories and general members of the research community at a level commensurate with the budget limitations. The resident archive should include software tools with user documentation needed for accessing and displaying the resident data.

The RA should maintain reserves such that if the RA maintenance award is not renewed or is subsumed under another RA structure, the RA would transfer the data to the other

Heliophysics Science Data Management Policy

RA or ensure permanent archiving by arrangement with the NSSDC. The RA proposal should include a plan for such transfer to a another archive in a manner that will still allow basic data access.

RAs are not intended to generate significant upgrades to the data sets, reprocess data, upgrade data processing algorithms, or provide new data products derived from the resident data. These types of post-mission data activities need to be funded from other sources. On the other hand, maintaining a resident archive could include “loading” newly derived data products into the archive with appropriate changes to metadata, documentation, web interfaces, etc.

Heliophysics Science Data Management Policy

Heliophysics Mission Data Lifecycle

Event	Launch	End of Prime	End of New Data	End of Mission	Final Archiving
Data Actions	Data system production	Reduction, PI serving, product improvement	Final data file prep., PI Serving	Serving; PI help	Preserving Serving
Archive Status		Active Archive (mission-based; others possible in all phases)		Resident Archive	Permanent Archive
Data Mgmt Document	PDMP Prep. and approval	PDMP use; MAP preparation and use in the Extended mission		RA plan	Archive Plan
General Serving		Virtual Observatories and Services			
Management	Mission (Program/Project)		Mission (Prog./Proj.) (with Senior Review)	NSSDC	

Figure 1: HP Mission Data Lifecycle

Heliophysics Science Data Management Policy

Acronym list

AA Active Archive (serves data during the active mission lifetime)

API Application Programmer Interface

CDF Common Data Format (self-documenting format commonly used in space physics)

DE Data Environment

FITS Flexible Image Transport System (self-documenting format commonly used in solar physics and astronomy)

HDF Hierarchical Data Format (self-documenting format commonly used in earth science)

HP Heliophysics

HPDCWG Heliophysics Data and Computing Working Group

IVOA International Virtual Observatory Alliance (astronomy)

MAP Mission Archive Plan (post PDMP plan for data product generation, etc.)

netCDF network Common Data Format (self-documenting format used in some areas of space and earth science)

NSSDC National Space Science Data Center (NASA's permanent archiving facility)

NVO National Virtual Observatory (astronomy)

PDMP Project Data Management Plan

PDS Planetary Data System

RA Resident Archive (serves data after the end of a mission)

SARC Science Archive Research Center (astronomical NASA data center)

SDAC Solar Data Analysis Center (NASA solar data center)

SPASE Space Physics Archive, Search, and Extract (provider of a community Data Model)

Heliophysics Science Data Management Policy

SPDF Space Physics Data Facility (NASA Data Center for space physics)

VO Virtual Observatory

VxO VO for the “x” community